



Lee, S., Ding, F., Manby, F., & Miller, T. (2019). Analytical gradients for projection-based wavefunction-in-DFT embedding. *Journal of Chemical Physics*, 151, [064112 (2019)].  
<https://doi.org/10.1063/1.5109882>

Peer reviewed version

Link to published version (if available):  
[10.1063/1.5109882](https://doi.org/10.1063/1.5109882)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via AIP at <https://aip.scitation.org/doi/full/10.1063/1.5109882>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Analytical Gradients for Projection-Based Wavefunction-in-DFT Embedding

Sebastian J. R. Lee,<sup>1</sup> Feizhi Ding,<sup>1</sup> Frederick R. Manby,<sup>2</sup> and Thomas F. Miller III<sup>1, a)</sup>

<sup>1)</sup>*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States*

<sup>2)</sup>*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom*

(Dated: 3 July 2019)

Projection-based embedding provides a simple, robust, and accurate approach for describing a small part of a chemical system at the level of a correlated wavefunction method while the remainder of the system is described at the level of density functional theory. Here, we present the derivation, implementation, and numerical demonstration of analytical nuclear gradients for projection-based wavefunction-in-density functional theory (WF-in-DFT) embedding. The gradients are formulated in the Lagrangian framework to enforce orthogonality, localization, and Brillouin constraints on the molecular orbitals. An important aspect of the gradient theory is that WF contributions to the total WF-in-DFT gradient can be simply evaluated using existing WF gradient implementations without modification. Another simplifying aspect is that Kohn-Sham (KS) DFT contributions to the projection-based embedding gradient do not require knowledge of the WF calculation beyond the relaxed WF density. Projection-based WF-in-DFT embedding gradients are thus easily generalized to any combination of WF and KS-DFT methods. We provide numerical demonstration of the method for several applications, including calculation of a minimum energy pathway for a hydride transfer in a cobalt-based molecular catalyst using the nudged-elastic-band method at the CCSD-in-DFT level of theory, which reveals large differences from the transition state geometry predicted using DFT.

## I. INTRODUCTION

The theoretical description of many chemical processes demands accurate, *ab initio* electronic structure theories. However, the study of complex reactive processes, including those arising in inorganic and enzyme catalysis, gives rise to the need for a compromise between accuracy and the ability to complete the computation in a reasonable amount of time. For systems in which the complicated chemical rearrangements (e.g. bond breaking and forming) occurs in a local spatial region, an effective strategy for balancing accuracy and computational cost is to employ one of various multiscale embedding strategies.<sup>1–25</sup> Generally, embedding methodologies hinge on the condition that a system can be efficiently partitioned into a local subsystem that demands a high-level treatment and an environment that can be treated with a lower (and computationally less expensive) level of theory.

The current paper focuses on projection-based embedding,<sup>9,26</sup> a DFT-based embedding theory in which subsystem partitioning is performed in terms of localized Kohn-Sham (KS) molecular orbitals (LMOs). The method describes subsystem interactions at the level of KS and allows for the partitioning of the subsystems across covalent and even conjugated bonds, and it enables the use of relatively small subsystem sizes for an embedded WF description. A recent review of projection-based WF-in-DFT embedding is available in Ref. 26.

Projection-based embedding has proven to be a useful tool in a wide range of chemical contexts including transition-metal complexes,<sup>19,20,24,27</sup> protein ac-

tive sites,<sup>21,23</sup> excited states<sup>28–30</sup> and condensed phase systems,<sup>16</sup> among others.<sup>31–37</sup> The development of analytical nuclear gradients for projection-based embedding will expand its applicability to include geometry optimization, transition state searches, and potentially *ab initio* molecular dynamics. Analytical nuclear gradients already exist for a number of other embedding methodologies, including the incremental molecular fragmentation method,<sup>38</sup> fragment molecular orbital method,<sup>39–41</sup> QM/MM,<sup>42–44</sup> ONIOM,<sup>45–48</sup> EMFT,<sup>17,49</sup> frozen density embedding,<sup>50–53</sup> and subsystem DFT.<sup>54–56</sup> However, the projection-based approach provides a number of advantages for WF-in-DFT embedding calculations and leads to a distinct analytical gradient theory, which we derive and numerically demonstrate in several applications.

In section II A we outline projection-based WF-in-DFT embedding and in section II B we provide the derivation of its analytical nuclear gradients. Section IV numerically validates the analytical nuclear gradient theory and its implementation in Molpro<sup>57</sup> via comparison with finite difference calculations, as well as presenting results for optimizing geometries in benchmark systems and the calculation of a minimum energy profile for an organometallic reaction using the nudged-elastic-band (NEB) method. We additionally provide the analytical nuclear gradient theory for WF-in-DFT embedding with atomic orbital (AO) truncation<sup>18</sup> in Appendices C and D.

---

<sup>a)</sup> tfm@caltech.edu

## II. PROJECTION-BASED EMBEDDING ANALYTICAL NUCLEAR GRADIENTS

### A. Projection-based Embedding Energy Theory

Projection-based WF-in-DFT embedding relies on the partitioning the LMOs of a system into two subsystems. Subsystem A contains the LMOs that are treated using the WF method and subsystem B contains the remaining LMOs that are treated using KS. This WF-in-DFT procedure is accomplished by first performing a KS calculation on the full system to obtain a set of KS MOs. The occupied KS MOs are then localized and partitioned into subsystems A and B. Finally, subsystem A is treated using the WF method in the presence of the embedding potential created by the frozen LMOs of subsystem B. Note that the cost of the KS calculation on the full system is typically negligible in comparison to the subsystem WF calculation. This results in our working equation for projection-based WF-in-DFT embedding,<sup>9</sup>

$$E_{\text{WF-in-DFT}}[\tilde{\Psi}^A; \gamma^A, \gamma^B] = E_{\text{WF}}[\tilde{\Psi}^A] + \text{tr}[(\tilde{\mathbf{d}}^A - \gamma^A)\mathbf{v}_{\text{emb}}[\gamma^A, \gamma^B]] + E_{\text{DFT}}[\gamma^A + \gamma^B] - E_{\text{DFT}}[\gamma^A] + \mu \text{tr}[\tilde{\mathbf{d}}^A \mathbf{P}^B], \quad (1)$$

where  $\tilde{\Psi}^A$  and  $E_{\text{WF}}[\tilde{\Psi}^A]$  are the WF and energy of subsystem A,  $\tilde{\mathbf{d}}^A$  is the subsystem A one-particle reduced density matrix that corresponds to  $\tilde{\Psi}^A$ ,  $E_{\text{DFT}}$  is the KS energy, and  $\gamma^A$  and  $\gamma^B$  are respectively the KS subsystem A and B one-particle densities that equal the full system KS density,  $\gamma$ , when summed together. Throughout, we shall use a tilde to indicate quantities that have been calculated using the WF method. The embedding potential,  $\mathbf{v}_{\text{emb}}$ , is defined as

$$\mathbf{v}_{\text{emb}}[\gamma^A, \gamma^B] = \mathbf{g}[\gamma^A + \gamma^B] - \mathbf{g}[\gamma^A], \quad (2)$$

where  $\mathbf{g}$  includes all KS two-electron terms,

$$(\mathbf{g}[\gamma])_{\kappa\nu} = \sum_{\lambda\sigma} \gamma_{\lambda\sigma} \left( (\kappa\nu|\lambda\sigma) - \frac{1}{2} x_f (\kappa\lambda|\nu\sigma) \right) + (\mathbf{v}_{\text{xc}}[\gamma])_{\kappa\nu}, \quad (3)$$

and where  $\kappa, \nu, \lambda$  and  $\sigma$  label atomic orbital basis functions,  $(\kappa\nu|\lambda\sigma)$  are the two-electron repulsion integrals,  $x_f$  is the fraction of exact exchange and  $\mathbf{v}_{\text{xc}}$  is the exchange-correlation (XC) potential matrix. The level-shift operator,  $\mu\mathbf{P}^B$ , is given by

$$\mu\mathbf{P}^B = \mu\mathbf{S}\gamma^B\mathbf{S}, \quad (4)$$

where  $\mathbf{S}$  is the overlap matrix. In the limit of  $\mu \rightarrow \infty$ , the LMOs that make up subsystems A and B are enforced to be exactly orthogonal, eliminating the non-additive kinetic energy present in other embedding frameworks.<sup>58,59</sup> In practice, finite values of  $\mu$  in the range of  $10^4$  hartree to

$10^7$  hartree are found to provide accurate results regardless of chemical system.<sup>9</sup> If greater accuracy is needed, a perturbative correction outlined in Ref. 9 can be added to the WF-in-DFT energy expression to account for the finiteness of  $\mu$ , but in practice, this correction is found to contribute negligibly to the total energy and is thus neglected here.

Projection-based embedding can also be used for DFT-in-DFT embedding via a simplified version of Eq. 1. The working equation for projection-based DFT-in-DFT embedding is<sup>9</sup>

$$E_{\text{DFT-in-DFT}}[\tilde{\gamma}^A; \gamma^A, \gamma^B] = E_{\text{DFT}}[\tilde{\gamma}^A] + \text{tr}[(\tilde{\gamma}^A - \gamma^A)\mathbf{v}_{\text{emb}}[\gamma^A, \gamma^B]] + E_{\text{DFT}}[\gamma^A + \gamma^B] - E_{\text{DFT}}[\gamma^A] + \mu \text{tr}[\tilde{\gamma}^A \mathbf{P}^B]. \quad (5)$$

The only differences between WF-in-DFT and DFT-in-DFT embedding is that the first term on the RHS of Eq. 1 is replaced with the KS energy on subsystem A,  $E_{\text{DFT}}[\tilde{\gamma}^A]$ , and in the second and last terms  $\tilde{\mathbf{d}}^A$  is reduced to the subsystem A KS density matrix,  $\tilde{\gamma}^A$ .

### B. Projection-based WF-in-DFT Embedding Gradient Theory

Since projection-based embedding is a non-variational theory, its analytical gradient is conveniently derived using a Lagrangian approach. We first construct a Lagrangian based on the projection-based WF-in-DFT energy. We then minimize the Lagrangian with respect to the variational parameters in the embedding energy, which include the subsystem A WF and the LMO coefficients. Then we show how to solve for each of the Lagrange multipliers and provide the working equation for the gradient of the total energy.

For consistency in notation, the MO coefficient matrix  $\mathbf{C}$  refers to the entire set of KS MOs (occupied and virtual). The submatrix of  $\mathbf{C}$  that refers to the (occupied) LMOs is denoted as  $\mathbf{L}$  with column indices  $i, j, k, l$ . The submatrix of  $\mathbf{C}$  that refers to the canonical virtual space is denoted as  $\mathbf{C}_v$  with column indices  $a, b, c, d$ . The indices  $m, n, p, q$  are used to index generic molecular orbitals.

### C. Total Energy Lagrangian

We now derive the total energy Lagrangian for projection-based WF-in-DFT embedding. Where appropriate we will provide WF method specific examples (e.g. MP2) of general terms outlined in the equations. The

WF-in-DFT Lagrangian is

$$\begin{aligned} \mathcal{L} [\mathbf{C}, \tilde{\Psi}^A, \mathbf{A}, \mathbf{x}, \mathbf{z}^{\text{loc}}, \mathbf{z}] = & \\ E_{\text{WF-in-DFT}} [\tilde{\Psi}^A; \gamma^A, \gamma^B] + \sum_s \Lambda_s^{\text{WF},A} c_s & \\ + \sum_{pq} x_{pq} (\mathbf{C}^\dagger \mathbf{S} \mathbf{C} - \mathbf{1})_{pq} + \sum_{i>j} z_{ij}^{\text{loc}} r_{ij} & \\ + \sum_{ai} z_{ai} (\mathbf{F} [\gamma^A + \gamma^B])_{ai}. & \end{aligned} \quad (6)$$

The first term on the right hand side (RHS) of Eq. 6 is the projection-based WF-in-DFT embedding energy described by Eq. 1. The second term on the RHS of Eq. 6 contains any constraints,  $c_s$ , and the corresponding Lagrange multipliers,  $\Lambda_s^{\text{WF},A}$ , that arise from ensuring that the Lagrangian is variational with respect to parameters in the WF method. The third term on the RHS constrains the KS MOs,  $\mathbf{C}$ , to be orthonormal, which accounts for the basis set being atom centered; this term is commonly referred to as the Pulay force<sup>60</sup> and arises from the atomic orbital basis set being atom centered. The localization conditions,  $r_{ij} = 0$ , take into account how the KS MOs are localized before being selected for subsystems A and B. This is important because the LMOs will have a different dependence on nuclear perturbation than canonical MOs. In this work, we use Pipek-Mezey localization<sup>61</sup> to obtain LMOs. Generalization to other localization methods (e.g. Boys<sup>62</sup> and intrinsic bond orbitals<sup>63</sup>) is straightforward. The localization conditions for Pipek-Mezey are

$$r_{ij} = \sum_C (S_{ii}^C - S_{jj}^C) S_{ij}^C = 0 \quad \text{for all } i > j, \quad (7)$$

where  $C$  corresponds to an atom in the molecule. The matrices  $S^C$  are defined as

$$S_{kl}^C = \sum_{\alpha \in C} \sum_{\beta} (L_{\alpha,k} S_{\alpha\beta} L_{\beta,l} + L_{\alpha,l} S_{\alpha\beta} L_{\beta,k}), \quad (8)$$

where the summation over  $\alpha$  is restricted to basis functions at atom  $C$ . The Brillouin conditions,  $(\mathbf{F} [\gamma^A + \gamma^B])_{ai} = 0$ , reflect how the KS MOs are optimized before being used to construct subsystems A and B. The Brillouin conditions are only needed because subsystem B is frozen at the KS level of theory. However, due to the non-additivity of the XC potential, the Lagrange multipliers,  $\mathbf{z}$ , span the full virtual-occupied space.

The type and number of constraints applied to the WF method depend on the chosen method. For example, if the WF method is MP2 then the constraints are

$$\begin{aligned} \sum_s \Lambda_s^{\text{MP2},A} c_s = \sum_{pq} \tilde{x}_{pq} (\tilde{\mathbf{C}}^{A\dagger} \mathbf{S} \tilde{\mathbf{C}}^A - \mathbf{1})_{pq} & \\ + \sum_{ai} \tilde{z}_{ai} (\mathbf{F}^A)_{ai} |_{i \in A}, & \end{aligned} \quad (9)$$

where the first term on the RHS of Eq. 9 constrains the Hartree-Fock MOs,  $\tilde{\mathbf{C}}^A$ , to be orthonormal, the condition

$i \in A$  restricts the sum to occupied MOs in subsystem A, and the second term on the RHS are the Brillouin conditions using the embedded Fock matrix,  $\mathbf{F}^A$ . The embedded Fock matrix is defined as<sup>9</sup>

$$\mathbf{F}^A = \mathbf{h} + \mathbf{g} [\tilde{\gamma}^A] + \mathbf{v}_{\text{emb}} [\gamma^A, \gamma^B] + \mu \mathbf{P}^B, \quad (10)$$

where  $\mathbf{h}$  is the standard one-electron Hamiltonian,  $\mathbf{g}$  includes all of the usual HF two-electron terms and  $\tilde{\gamma}^A$  is the subsystem A HF one-particle density. These constraints also arise in the derivation of the MP2 analytical nuclear gradient.<sup>64</sup>

For the projection-based WF-in-DFT energy to equal the Lagrangian, the Lagrangian must be minimized with respect to all of its parameters, including  $\tilde{\Psi}^A$ ,  $\mathbf{C}$ , and all of the Lagrange multipliers.

#### D. Minimizing the Lagrangian with respect to the variational parameters of the WF method

Upon minimizing the WF-in-DFT Lagrangian with respect to  $\tilde{\Psi}^A$ , only terms associated with the first two terms on the RHS of Eq. 6 survive, all of which are familiar from the WF Lagrangian for the corresponding WF gradient theories.

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \tilde{\Psi}^A} = \frac{\partial E_{\text{WF}} [\tilde{\Psi}^A]}{\partial \tilde{\Psi}^A} + \text{tr} \left[ \frac{\partial \tilde{\mathbf{d}}^A}{\partial \tilde{\Psi}^A} \mathbf{v}_{\text{emb}} [\gamma^A, \gamma^B] \right] & \\ + \mu \text{tr} \left[ \frac{\partial \tilde{\mathbf{d}}^A}{\partial \tilde{\Psi}^A} \mathbf{P}^B \right] + \frac{\partial}{\partial \tilde{\Psi}^A} \sum_s \Lambda_s^{\text{WF},A} c_s = 0 & \end{aligned} \quad (11)$$

Since the embedding potential is independent of  $\tilde{\Psi}^A$ , the Z-vector coupled perturbed Hartree-Fock (Z-CPHF) equations of any post-HF method are only impacted through the eigenvalues of the subsystem A HF WF. Therefore, the solutions for the WF Lagrange multipliers (e.g.  $\tilde{\mathbf{x}}$  and  $\tilde{\mathbf{z}}$  for MP2 in Eq. 9) are obtained using the standard implementation of the WF gradient no matter what KS method is selected to describe subsystem B. However, if an alternative embedding potential is used that depends on the subsystem A WF, such as the Huzinaga constraint (i.e. Ref. 65), then the formulation of the WF gradient is changed; the Z-CPHF equations for a general WF method would need to be modified to include the contributions from the derivative of the embedding potential with respect to the subsystem A WF,  $\tilde{\Psi}^A$ .

#### E. Minimizing the Lagrangian with respect to the MO coefficients

The remaining Lagrange multipliers,  $\mathbf{z}^{\text{loc}}$ ,  $\mathbf{x}$ , and  $\mathbf{z}$  in Eq. 6, are determined by minimizing the WF-in-DFT Lagrangian with respect to the variational parameters of the KS method, namely the MO coefficients,  $\mathbf{C}$ . Differentiation of the Lagrangian with respect to these parameters

yields

$$\sum_{\mu} C_{\mu,p} \frac{\partial \mathcal{L}}{\partial C_{\mu,q}} = E_{pq} + (\mathbf{a}[\mathbf{z}^{\text{loc}}])_{pi} + (\mathbf{D}[\mathbf{z}])_{pq} + 2x_{pq} = 0, \quad (12)$$

where

$$E_{pq} = \sum_{\mu} C_{\mu,p} \left( \frac{\partial E_{\text{WF-in-DFT}}[\tilde{\Psi}^A; \gamma^A, \gamma^B]}{\partial C_{\mu,q}} + \frac{\partial}{\partial C_{\mu,q}} \sum_s \Lambda_s^{\text{WF,A}} c_s \right), \quad (13)$$

$$\begin{aligned} (\mathbf{a}[\mathbf{z}^{\text{loc}}])_{pi} &= \sum_{\mu} C_{\mu,p} \left( \sum_{k>l} z_{kl}^{\text{loc}} \frac{\partial r_{kl}}{\partial C_{\mu,q}} \right) \\ &= \sum_{k>l} \mathcal{B}_{pi,kl} z_{kl}^{\text{loc}}, \end{aligned} \quad (14)$$

$$\begin{aligned} (\mathbf{D}[\mathbf{z}])_{pq} &= \sum_{\mu} C_{\mu,p} \left( \sum_{ak} z_{ak} \frac{\partial (\mathbf{F}[\gamma^A + \gamma^B])_{ak}}{\partial C_{\mu,q}} \right) \\ &= \sum_{ak} \mathcal{D}_{pq,ak} z_{ak} \\ &= (\mathbf{F}[\gamma^A + \gamma^B] \mathbf{z})_{pq} \Big|_{q \in \text{occ}} \\ &\quad + (\mathbf{F}[\gamma^A + \gamma^B] \mathbf{z}^\dagger)_{pq} \Big|_{q \in \text{vir}} + 2(\mathbf{V}[\bar{\mathbf{z}}])_{pq} \Big|_{q \in \text{occ}}, \end{aligned} \quad (15)$$

and

$$2x_{pq} = \sum_{\mu} C_{\mu,p} \left( \sum_{mn} x_{mn} \frac{\partial S_{mn}}{\partial C_{\mu,q}} \right). \quad (16)$$

The 4-dimensional tensors,  $\mathcal{B}$  and  $\mathcal{D}$ , are expanded in Appendices A and B, respectively,  $\bar{\mathbf{z}}$  corresponds to  $\mathbf{z} + \mathbf{z}^\dagger$ , and  $\mathbf{V}[\bar{\mathbf{z}}]$  includes all two-electron terms of the generalized Fock matrix and is shown explicitly in Appendix B. Since the embedded Fock matrix,  $\mathbf{F}^A$ , contains the embedding potential,  $\mathbf{v}_{\text{emb}}$ , its derivative with respect to the MO coefficients,  $\mathbf{C}$ , is nonzero resulting in the WF relaxed density being needed to construct  $\mathbf{E}$  in Eq. 13, which is explicitly shown in Appendix B. Therefore, the subsystem A WF gradient only affects the embedding contributions to the gradient through the WF relaxed density.

We now show that solving for the Lagrange multipliers leads to familiar coupled perturbed equations. Combining the stationary conditions described by Eq. 12 with the auxiliary conditions  $\mathbf{x} = \mathbf{x}^\dagger$  yields the linear Z-vector equations<sup>64</sup>

$$(1 - \mathcal{P}_{pq})(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{a}[\mathbf{z}^{\text{loc}}])_{pq} = 0, \quad (17)$$

where  $\mathcal{P}_{pq}$  permutes the indices  $p$  and  $q$ , which is used to solve for  $\mathbf{z}$  and  $\mathbf{z}^{\text{loc}}$ . The matrix  $\mathbf{x}$  is then obtained as

$$x_{pq} = -\frac{1}{4}(1 + \mathcal{P}_{pq})(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{a}[\mathbf{z}^{\text{loc}}])_{pq}. \quad (18)$$

The Lagrange multipliers  $\mathbf{z}^{\text{loc}}$  pertain to the occupied-occupied MO space; considering only the occupied-occupied part of Eq. 17 yields

$$(1 - \mathcal{P}_{ij})(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{a}[\mathbf{z}^{\text{loc}}])_{ij} = 0. \quad (19)$$

Using the Brillouin conditions and the knowledge that  $z_{ab} = z_{ij} = z_{ia} = 0$ , Eq. 19 can be further simplified by showing that

$$(1 - \mathcal{P}_{ij})(\mathbf{D}[\mathbf{z}])_{ij} = 0. \quad (20)$$

The solutions,  $\mathbf{z}^{\text{loc}}$ , are thus independent of  $\mathbf{z}$ , such that Eq. 19 reduces to

$$E_{ij} - E_{ji} + \sum_{k>l} (\mathcal{B}_{ij,kl} - \mathcal{B}_{ji,kl}) z_{kl}^{\text{loc}} = 0. \quad (21)$$

These are the Z-vector coupled perturbed localization (Z-CPL) equations, which are used to solve for  $\mathbf{z}^{\text{loc}}$ . Subsequently,  $\mathbf{a}[\mathbf{z}^{\text{loc}}]$  can be computed according to Eq. 14.

The Lagrange multipliers  $\mathbf{z}$  pertain to the virtual-occupied MO space; considering only the virtual-occupied part of Eq. 17 yields

$$(1 - \mathcal{P}_{ai})(\mathbf{E} + \mathbf{D}[\mathbf{z}] + \mathbf{a}[\mathbf{z}^{\text{loc}}])_{ai} = 0, \quad (22)$$

which further simplifies to

$$\begin{aligned} (\mathbf{E} + \mathbf{a}[\mathbf{z}^{\text{loc}}] + \mathbf{F}[\gamma^A + \gamma^B] \mathbf{z} - \mathbf{z} \mathbf{F}[\gamma^A + \gamma^B] \\ + 2\mathbf{V}[\bar{\mathbf{z}}])_{ai} = 0. \end{aligned} \quad (23)$$

These are the Z-vector coupled perturbed Kohn-Sham (Z-CPKS) equations. Having solved the Z-CPL and Z-CPKS equations, the remaining Lagrangian multipliers associated with the orthogonality constraints,  $\mathbf{x}$ , can be obtained from Eq. 18.

## F. Gradient of the Total Energy

Once the Lagrangian is minimized with respect to all variational parameters, the gradient of the total energy takes the form

$$\begin{aligned} \frac{dE_{\text{WF-in-DFT}}}{dq} &= \frac{d\mathcal{L}}{dq} = \frac{\partial \mathcal{L}}{\partial q} + \frac{\partial \mathcal{L}}{\partial \tilde{\Psi}^A} \frac{\partial \tilde{\Psi}^A}{\partial q} + \frac{\partial \mathcal{L}}{\partial \mathbf{C}} \frac{\partial \mathbf{C}}{\partial q} \\ &= \frac{\partial \mathcal{L}}{\partial q}. \end{aligned} \quad (24)$$

Since the Lagrangian is minimized with respect to the subsystem A WF and the KS LMO coefficients, calculation of the WF and KS LMO responses to nuclear perturbation,  $\partial \tilde{\Psi}^A / \partial q$  and  $\partial \mathbf{C} / \partial q$  respectively, are avoided.

This yields the following expression for the gradient,

$$\begin{aligned}
E_{\text{WF-in-DFT}}^{(q)} &= E_{\text{WF}}^{(q)} \left[ \tilde{\Psi}^A \right] + \sum_{\lambda\nu} \tilde{\mathbf{d}}_{\lambda\nu}^A (\mathbf{v}_{\text{emb}}^{(q)})_{\lambda\nu} \\
&+ \mu \sum_{\lambda\nu} \tilde{\mathbf{d}}_{\lambda\nu}^A (\mathbf{P}^{B,(q)})_{\lambda\nu} + \sum_s \Lambda_s^{\text{WF},A} c_s^{(q)} \\
&- E_{\text{DFT}}^{(q)} [\gamma^A] + E_{\text{DFT}}^{(q)} [\gamma^A + \gamma^B] - \sum_{\lambda\nu} \gamma_{\lambda\nu}^A (\mathbf{v}_{\text{emb}}^{(q)})_{\lambda\nu} \\
&+ \sum_{ij} z_{ij}^{\text{loc}} r_{ij}^{(q)} + \sum_{ai} z_{ai} (\mathbf{F} [\gamma^A + \gamma^B])_{ai}^{(q)} + \sum_{mn} x_{mn} S_{mn}^{(q)},
\end{aligned} \tag{25}$$

where the superscript  $(q)$  denotes the explicit derivative of the quantity with respect to a nuclear coordinate. Eq. 25 can be further simplified by folding  $\sum_s \Lambda_s^{\text{WF},A} c_s^{(q)}$  into the first three terms on the RHS of Eq. 25, yielding

$$\begin{aligned}
E_{\text{WF-in-DFT}}^{(q)} &= E_{\text{WF}}^q \left[ \tilde{\Psi}^A \right] + \sum_{\lambda\nu} (\tilde{\mathbf{d}}_{\text{rel}}^A)_{\lambda\nu} (\mathbf{v}_{\text{emb}}^{(q)})_{\lambda\nu} \\
&+ \mu \sum_{\lambda\nu} (\tilde{\mathbf{d}}_{\text{rel}}^A)_{\lambda\nu} (\mathbf{P}^{B,(q)})_{\lambda\nu} \\
&- E_{\text{DFT}}^{(q)} [\gamma^A] + E_{\text{DFT}}^{(q)} [\gamma^A + \gamma^B] - \sum_{\lambda\nu} \gamma_{\lambda\nu}^A (\mathbf{v}_{\text{emb}}^{(q)})_{\lambda\nu} \\
&+ \sum_{ij} z_{ij}^{\text{loc}} r_{ij}^{(q)} + \sum_{ai} z_{ai} (\mathbf{F} [\gamma^A + \gamma^B])_{ai}^{(q)} + \sum_{mn} x_{mn} S_{mn}^{(q)}.
\end{aligned} \tag{26}$$

Here,  $E_{\text{WF}}^q [\tilde{\Psi}^A]$  denotes the total derivative of the subsystem A WF energy with respect to nuclear coordinate, which can be directly calculated using existing WF gradient implementations, and  $\tilde{\mathbf{d}}_{\text{rel}}^A$  is the WF-relaxed density for subsystem A. For example, the MP2-relaxed density is

$$\tilde{\mathbf{d}}_{\text{rel}}^A = \tilde{\mathbf{d}}^A + \tilde{\mathbf{C}}^A \tilde{\mathbf{z}} \tilde{\mathbf{C}}^{A,\dagger} = \tilde{\gamma}^A + \mathbf{d}^{(2)} + \tilde{\mathbf{C}}^A \tilde{\mathbf{z}} \tilde{\mathbf{C}}^{A,\dagger}, \tag{27}$$

which contains the subsystem A Hartree-Fock density,  $\tilde{\gamma}^A$ , the MP2 density matrix,  $\mathbf{d}^{(2)}$ , and the solutions of the subsystem A Brillouin conditions,  $\tilde{\mathbf{C}}^A \tilde{\mathbf{z}} \tilde{\mathbf{C}}^{A,\dagger}$ . Eq. 26 can be expressed in terms of the WF gradient on subsystem A and the derivative AO integrals, yielding our final expression for the projection-based WF-in-DFT analytical gradient,

$$\begin{aligned}
E_{\text{WF-in-DFT}}^{(q)} &= E_{\text{WF}}^q \left[ \tilde{\Psi}^A \right] \\
&+ \text{tr} \left[ \mathbf{d}_a \mathbf{h}^{(q)} \right] + \text{tr} \left[ \mathbf{X} \mathbf{S}^{(q)} \right] + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} D_{\mu\nu\lambda\sigma} (\mu\nu|\lambda\sigma)^{(q)} \\
&+ E_{\text{xc}}^{(q)} [\gamma^A + \gamma^B] - E_{\text{xc}}^{(q)} [\gamma^A] \\
&+ \text{tr} \left[ \left( \tilde{\mathbf{d}}_{\text{rel}}^A - \gamma^A \right) \left( \mathbf{v}_{\text{xc}}^{(q)} [\gamma^A + \gamma^B] - \mathbf{v}_{\text{xc}}^{(q)} [\gamma^A] \right) \right].
\end{aligned} \tag{28}$$

The effective one-particle density  $\mathbf{d}_a$  and effective two-particle density  $\mathbf{D}$  are defined

$$\mathbf{d}_a = \gamma^B + \mathbf{C} \mathbf{z} \mathbf{C}^\dagger, \tag{29}$$

and

$$\begin{aligned}
D_{\mu\nu\lambda\sigma} &= (\gamma^A + \gamma^B)_{\mu\nu} (\mathbf{d}_b)_{\lambda\sigma} - \gamma_{\mu\nu}^A (\mathbf{d}_c)_{\lambda\sigma} \\
&- \frac{1}{2} x_f \left( (\gamma^A + \gamma^B)_{\mu\lambda} (\mathbf{d}_b)_{\nu\sigma} - \gamma_{\mu\lambda}^A (\mathbf{d}_c)_{\nu\sigma} \right).
\end{aligned} \tag{30}$$

The effective one-particle densities  $\mathbf{d}_b$  and  $\mathbf{d}_c$  are defined

$$\mathbf{d}_b = \gamma^A + \gamma^B + 2\mathbf{C} \mathbf{z} \mathbf{C}^\dagger + 2\tilde{\mathbf{d}}_{\text{rel}}^A - 2\gamma^A, \tag{31}$$

and

$$\mathbf{d}_c = -\gamma^A + 2\tilde{\mathbf{d}}_{\text{rel}}^A. \tag{32}$$

The matrix  $\mathbf{X}$  is defined

$$\begin{aligned}
\mathbf{X} &= \mathbf{C} \mathbf{x} \mathbf{C}^\dagger + \sum_{i>j} \frac{\partial r_{ij}}{\partial S_{\mu\nu}} z_{ij}^{\text{loc}} \\
&= \mathbf{X}^{\text{loc}} - \frac{1}{2} \mathbf{L} (\mathbf{E} + 2\mathbf{V}[\bar{\mathbf{z}}]) \mathbf{L}^\dagger \\
&- \frac{1}{2} \left( \mathbf{C}_v (\mathbf{z} \mathbf{F}) \mathbf{L}^\dagger + (\mathbf{C}_v (\mathbf{z} \mathbf{F}) \mathbf{L}^\dagger)^\dagger \right) \\
&+ \mu \left( \tilde{\mathbf{d}}_{\text{rel}}^A \mathbf{S} \gamma^B + \gamma^B \mathbf{S} \tilde{\mathbf{d}}_{\text{rel}}^A \right),
\end{aligned} \tag{33}$$

where

$$(\mathbf{X}^{\text{loc}})_{\mu\nu} = -\frac{1}{2} (\mathbf{L} \mathbf{a} [\mathbf{z}^{\text{loc}}] \mathbf{L}^\dagger)_{\mu\nu} + \sum_{i>j} \frac{\partial r_{ij}}{\partial S_{\mu\nu}} z_{ij}^{\text{loc}}. \tag{34}$$

The second term on the RHS of Eq. 34 is expanded in Appendix A.

The analytical nuclear gradient expression for projection-based DFT-in-DFT closely follows that for WF-in-DFT, with regard to evaluation of both the Lagrange multipliers (Eq. 12) and the final gradient (Eq. 28). To obtain the corresponding DFT-in-DFT expressions,  $\tilde{\mathbf{d}}_{\text{rel}}^A$  becomes the subsystem A KS density

$$\tilde{\mathbf{d}}_{\text{rel}}^A = \tilde{\gamma}^A, \tag{35}$$

which affects the evaluation of  $\mathbf{E}$  in Eq. 12 (expanded in Eq. B1) and the evaluation of the final gradient expression, Eq. 28. Additionally, the first term on the RHS of the final gradient expression, Eq. 28, is replaced with the subsystem A KS gradient,  $E_{\text{DFT}}^q [\tilde{\gamma}^A]$ .

### III. COMPUTATIONAL DETAILS

The implementation of projection-based WF-in-DFT embedding gradients is available in the 2019 general release of Molpro.<sup>57</sup> In all embedding calculations reported here, unless otherwise specified, the Pipek-Mezey localization method<sup>61</sup> is used with the core and occupied MOs localized together. The subsystem A region is chosen by including any LMOs with a net Mulliken population larger than 0.4 on the atoms associated with subsystem A, although more sophisticated partitioning algorithms

have been introduced.<sup>27</sup> A level-shift parameter of  $\mu = 10^6$  hartree is used for all embedding calculations. The perturbative correction to using a finite value of  $\mu$  in Eq. 5 is less than 20 microhartrees for the applications presented here and thus not included (accomplished by specifying the option `HF_COR = 0`). Throughout this work, all embedding calculations are described using the nomenclature “(WF method)-in-DFT/basis,” where the WF method describes subsystem A and the KS method describes subsystem B. For some embedding calculations a mixed-basis set is used and is denoted by “(WF method)-in-DFT/large-basis:small-basis,” where the large basis is used to describe subsystem A and the small basis is used to describe subsystem B.

All SCF calculations employ a tighter threshold than default for MO convergence by specifying the option `ORBITAL = 1 × 10-7` a.u. in Molpro. All KS calculations used in projection-based embedding are done without density fitting, employing the LDA,<sup>66,67</sup> PBE,<sup>68</sup> PBE0,<sup>69</sup> and LDAX functionals with the def2-TZVPP, def2-SVP, def2-ASVP,<sup>70,71</sup> cc-pVDZ,<sup>72</sup> and 6-31G<sup>73</sup> basis sets. Note that the def2-ASVP basis set used in Molpro is constructed by adding one set of even tempered diffuse functions to the def2-SVP basis set. The LDAX functional is constructed by including 50% exact exchange and reducing the weight of the DIRAC functional to 50% in the LDA functional. For the calculations in sections IV A and IV B 2, the XC functional is evaluated on a fixed-pruned grid with index 7 (Ref. 74). For the optimized geometries shown in section IV B, and the malondialdehyde calculations in section IV C the XC functional is evaluated on an adaptively generated quadrature grid that reproduces the energy of the Slater-Dirac functional to a specified threshold accuracy of  $10^{-10} E_h$ . All WF calculations are performed with the frozen-core approximation, without density fitting, employing the MP2,<sup>75</sup> CCSD,<sup>76,77</sup> and CCSD(T)<sup>78</sup> correlation treatments with the def2-TZVPP, def2-SVP, cc-pVDZ and 6-31G basis sets. Even though the density fitting approximation is not used for the WF methods in this study, density fitted gradients are available for the aforementioned WF methods.<sup>79,80</sup> The default values for integral screening were used in Molpro. For all Z-CPKS calculations an iterative subspace solver employing the Davidson algorithm<sup>81,82</sup> is used with a convergence threshold of  $1 \times 10^{-6}$  a.u. For all Z-CPHF calculations needed for the subsystem A WF gradient an iterative solver with a convergence threshold of  $1 \times 10^{-7}$  a.u. is used. Grid weight derivatives are included for all gradient calculations involving the XC functional and potential.

For all geometry optimizations the number of LMOs in subsystem A is kept unchanged throughout the optimization. A natural way of enforcing this in future work is to employ even-handed partitioning,<sup>27</sup> although this was not needed in the examples studied here; the default procedure based on net Mulliken population sufficed to keep subsystem A unchanged. All geometries are optimized using the translation-rotation-internal coordinate

system devised by Wang and Song,<sup>83</sup> which is available in the GeomeTRIC package.<sup>84</sup> Convergence parameters for the geometry optimizations follow the default parameters used by Molpro, namely that the maximum gradient value becomes less than  $3 \times 10^{-4}$  hartree/bohr and the energy change between adjacent steps becomes less than  $1 \times 10^{-6}$  hartree or the maximum component of the step displacement becomes less than  $3 \times 10^{-4}$  bohr. The maximum gradient value is evaluated in the Cartesian basis. All geometries are provided in the supporting information.

Nudged elastic band (NEB)<sup>85</sup> calculations are run using the implementation of the method in the atomic simulation environment (ASE) package.<sup>86</sup> All NEB calculations use Molpro forces which are provided through a Molpro calculator interface within the ASE package.

The intramolecular proton transfer of malondialdehyde is modeled with an NEB consisting of 15 images connected by springs with spring constants of  $0.1 \text{ eV/\AA}^2$ . The CCSD/def2-aSVP, CCSD-in-LDA/def2-aSVP and LDA/def2-aSVP optimized NEBs used the image dependent pair potential (IDPP) method<sup>87</sup> as the initial guess for the band with reactant and product geometries previously optimized at the corresponding level of theory. All NEB calculations for malondialdehyde are converged with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update of the Hessian and by enforcing that the maximum gradient value is less than  $0.01 \text{ eV/\AA}^2$ .

The intramolecular proton transfer of the organometallic cobalt complex is modeled with an NEB consisted of images connected by springs with spring constants of  $9 \text{ eV/\AA}^2$ . The PBE0/cc-pVDZ climbing image NEB,<sup>88</sup> consisting of 26 images, used the IDDP method as the initial guess for the band with the reactant geometry previously optimized. The CCSD-in-PBE0/cc-pVDZ NEB consisting of 23 images, used its optimized reactant and the climbing image NEB converged at the PBE0/cc-pVDZ level of theory as its initial guess. Since the product is spatially far away from the reactant, an intermediate geometry between the transition state and the product is used as the endpoint of the NEB. This intermediate geometry is determined by initially converging a NEB with an extra image such that the maximum force dropped below  $0.3 \text{ eV/\AA}^2$ . Then the second to last image is used as the new endpoint and a new NEB is converged. The PBE0/cc-pVDZ climbing image NEB is optimized using the FIRE<sup>89</sup> algorithm using a convergence criteria of  $0.05 \text{ eV/\AA}^2$  for the maximum gradient value. The projection-based WF-in-DFT embedding NEB is optimized at the CCSD-in-PBE0/cc-pVDZ level of theory using the FIRE algorithm with a convergence criteria of  $0.25 \text{ eV/\AA}^2$  for the maximum gradient value.

The CCSD-in-PBE0/cc-pVDZ calculations used for the NEB optimization are performed by specifying 51 occupied MOs to be in subsystem A using the `N_ORBITALS` option and by using AO truncation with a threshold of  $1 \times 10^{-3}$  a.u. An even-handed selection of AOs were used along the NEB by creating a union of the AOs that were

selected for each image by the truncation procedure to ensure that the NEB traversed a smooth potential energy surface. The final, reported energies of the WF-in-DFT NEB are performed using the PNO-LCCSD<sup>90</sup>/cc-pVDZ and PNO-LCCSD-in-PBE0/cc-pVDZ levels of theory. Both the PNO-LCCSD and PNO-LCCSD-in-PBE0 calculations are performed with density fitting using the cc-pVTZ/JKFIT<sup>91</sup> (the def2-TZVPP/JKFIT<sup>92</sup> basis set was used for cobalt since the cc-pVTZ/JKFIT basis set was not available) and the cc-pVTZ/MP2FIT<sup>93</sup> density fitting basis sets. Tighter domain approximations were employed for all PNO-LCCSD calculations by specifying the DOMOPT=TIGHT option. Additionally, the Boughton-Pulay completeness criterion was used for the selection of the primary projected atomic orbitals domain by specifying the option THRBP=1 and the Pipek-Mezey localization method was used. For the PNO-LCCSD-in-PBE0 calculations, AO truncation is not used, the core and valence DFT molecular orbitals are localized separately using the Pipek-Mezey localization method, and the subsystem A orbitals are selected using the default procedure based on the Mulliken population threshold.

All calculations using AO truncation<sup>18</sup> ensure that at least one AO is kept per atom (specified by option AO.PER.ATOM) to make evaluation of the integral derivative contributions from the one electron Hamiltonian simpler within Molpro. This adds a negligible amount of AO functions than would have been selected using only the density threshold parameter<sup>18</sup> for the systems studied in this paper. In all embedding geometry optimizations that employ AO truncation, the number of truncated AOs is fixed using the STOREAO option to ensure smoothness of the potential energy function. Upon convergence, the truncated AO list is reevaluated using the same density threshold parameter; if the number of kept AOs remains a subset of the original list of truncated AOs then the optimization is converged.

## IV. RESULTS AND DISCUSSION

### A. Comparison of Analytical and Numerical Gradients

The implementation of the projection-based WF-in-DFT analytical gradient is tested by comparison with the gradient evaluated by numerical finite difference for a distorted geometry of ethanol. The finite difference gradients are evaluated using a four-point central difference formula with a base step size of 0.01 bohr. The mean absolute error (MAE) between the analytical and finite difference gradients is reported for a range of embedding calculations in Table I. These results show that the analytical nuclear gradient for projection-based WF-in-DFT embedding is essentially numerically exact with respect to the gradients calculated by finite difference. Comparison of the results obtained using HF over the full system versus using LDA over the full system illustrate that some of the finite difference error comes from

TABLE I. Mean absolute error between the analytically and numerically determined embedding nuclear gradient for a distorted geometry of ethanol. The basis set 6-31G is used for all calculations. The distorted geometry of ethanol is provided in the supporting information.

Method	MAE (hartree/bohr)
HF	$5.00 \times 10^{-9}$
HF-in-HF	$4.61 \times 10^{-8}$
LDA	$1.48 \times 10^{-8}$
LDA-in-LDA	$7.23 \times 10^{-8}$
HF-in-LDA	$5.24 \times 10^{-8}$
MP2-in-LDA	$5.37 \times 10^{-8}$
CCSD-in-LDA	$5.36 \times 10^{-8}$
CCSD(T)-in-LDA	$5.26 \times 10^{-8}$
CCSD-in-LDA (AO) <sup>a</sup>	$3.48 \times 10^{-8}$
CCSD(T)-in-LDA (AO) <sup>a</sup>	$3.40 \times 10^{-8}$
CCSD(T)-in-PBE0	$5.26 \times 10^{-8}$
CCSD(T)-in-PBE0 (AO) <sup>a</sup>	$1.12 \times 10^{-7}$

<sup>a</sup>Calculations were performed with AO truncation with a density threshold of  $1 \times 10^{-1}$  a.u.

the DFT exchange-correlation grid. Comparison of the HF-in-HF results with full HF and of the LDA-in-LDA results with full LDA illustrate the modest effect of using a large-but-finite value for the level-shift operator in projection based embedding. These results confirm the correct implementation of projection-based WF-in-DFT analytical nuclear gradients.

### B. Optimized Geometries

#### 1. Ethanol

As a proof of concept, CCSD-in-LDA/6-31G analytical nuclear gradients are employed to determine the ground state geometry of ethanol, which is shown in Fig. 1. For this simple case, the O-H moiety is treated by CCSD and the remainder of the molecule is treated by LDA. Table II shows that the O-H bond length within subsystem A reproduces the CCSD predicted bond length of 0.979 Å and the remaining bonds within subsystem B reproduce the LDA predicted bond lengths. This indicates that the potential energy surface produced by projection-based embedding varies smoothly from CCSD-like interactions for subsystem A and LDA-like interactions for subsystem B. Interestingly, the C-O bond located at the boundary between subsystems A and B closely reproduces the LDA bond length and is not an interpolation between the CCSD and LDA bond lengths.



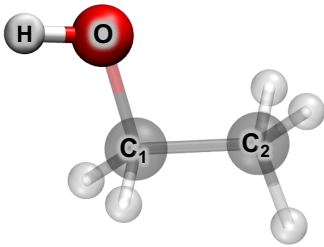


FIG. 1. Optimized geometry for ethanol using projection-based CCSD-in-LDA/6-31G. The solid atoms (O and H) are in subsystem A and the transparent atoms are in subsystem B.

TABLE II. Selected bond lengths and angles for ethanol (pictured in Fig.1) optimized at different levels of theory. Bond lengths are reported in units of Angstroms and angles are reported in units of degrees.

Method	r(O-H)	$\angle$ C <sub>1</sub> OH	r(C <sub>1</sub> -C <sub>2</sub> )	r(C <sub>1</sub> -O)
LDA/6-31G	0.988	110.4	1.503	1.439
CCSD-in-LDA/6-31G	0.979	110.7	1.506	1.435
CCSD/6-31G	0.979	110.6	1.532	1.475

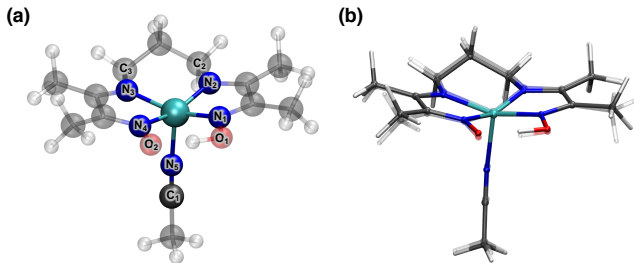


FIG. 2. (a) The optimized geometry for the cobalt-based organometallic complex performed with projection-based CCSD-in-LDAX/def2-TZVPP:def2-SVP with AO truncation. The solid atoms (Co, N<sub>1</sub>, N<sub>2</sub>, N<sub>3</sub>, N<sub>4</sub>, N<sub>5</sub>, and C<sub>1</sub>) are included in subsystem A and the transparent atoms are included in subsystem B. (b) The LDAX/def2-TZVPP:def2-SVP optimized geometry (transparent) and the projection-based CCSD-in-LDAX/def2-TZVPP:def2-SVP with AO truncation optimized geometry (solid).

## 2. Cobalt-based Organometallic Complex

As a demonstration of embedding gradients with AO truncation, the geometry of the cobalt-based organometallic complex, shown in Fig. 2, is optimized. Fig. 2a shows the CCSD-in-LDAX/def2-TZVPP:def2-SVP optimized structure of the cobalt complex where the solid atoms are included in subsystem A and the transparent atoms are included in subsystem B. In Fig. 2b the optimized structures evaluated at the CCSD-in-LDAX/def2-TZVPP:def2-SVP (solid) and the LDAX/def2-TZVPP:def2-SVP (transparent) levels of theory are overlaid. While only modest differences are

TABLE III. Selected bond lengths for the organometallic complex pictured in Fig.2 optimized at different levels of theory and their absolute difference ( $|\Delta|$ ). Bond lengths are reported in units of angstroms.

		LDAX	CCSD-in-LDAX	$ \Delta $
Sub A	r(Co-N <sub>1</sub> )	1.836	1.846	0.010
	r(Co-N <sub>2</sub> )	1.893	1.883	0.010
	r(Co-N <sub>3</sub> )	1.932	1.951	0.019
	r(Co-N <sub>4</sub> )	1.900	1.926	0.026
	r(Co-N <sub>5</sub> )	1.978	2.026	0.048
Boundary	r(N <sub>5</sub> -O <sub>1</sub> )	1.317	1.355	0.038
	r(N <sub>5</sub> -O <sub>2</sub> )	1.262	1.301	0.039
	r(C <sub>1</sub> -N <sub>5</sub> )	1.131	1.150	0.019
	r(N <sub>2</sub> -C <sub>2</sub> )	1.430	1.458	0.028
	r(N <sub>3</sub> -C <sub>3</sub> )	1.428	1.458	0.030
Sub B	r(O <sub>1</sub> -H)	1.025	1.030	0.005

seen in the overall structure, Table III shows that the optimized bond lengths do change between the two levels of theory, both for the region within subsystem A and at the subsystem boundary. This indicates that the WF method is capable of relaxing the atoms in subsystem A even when they are strongly coordinated with subsystem B. It is also seen that the bond lengths across the boundary of subsystems A and B also differ from the LDAX geometry since the bonds in question experience the effects of both the WF and KS methods. Finally, if a bond length associated with atoms in subsystem B is considered, such as the O<sub>1</sub>-H bond, it is found to closely match the LDAX predicted bond length.

## C. Malondialdehyde: Minimum Energy Reaction Pathway

The minimum energy reaction pathway for the proton transfer in malondialdehyde is determined using the NEB method. Fig. 3 shows that with minimal embedding (Fig. 3a) the CCSD-in-LDA/def2-aSVP reaction barrier, shown in Fig. 3b, is 4.85 kcal/mol which is within 1.5 kcal/mol of the CCSD/def2-aSVP reference reaction barrier of 6.12 kcal/mol. This is a vast improvement over the LDA/def2-aSVP result, which predicts an essentially barrierless reaction. In addition to correctly predicting the reaction barrier, Fig. 3c shows that the CCSD-in-LDA/def2-aSVP reaction pathway lies precisely on top of the CCSD/def2-aSVP pathway with only a small deviation in the basins. In contrast, Fig. 3d shows that the LDA/def2-aSVP reaction pathway and potential energy surface reveal errors in the location of the reactant and product basins, with the hydrogen-bond length vastly underestimated. This is consistent with the tendency of LDA to over stabilize hydrogen bonds.

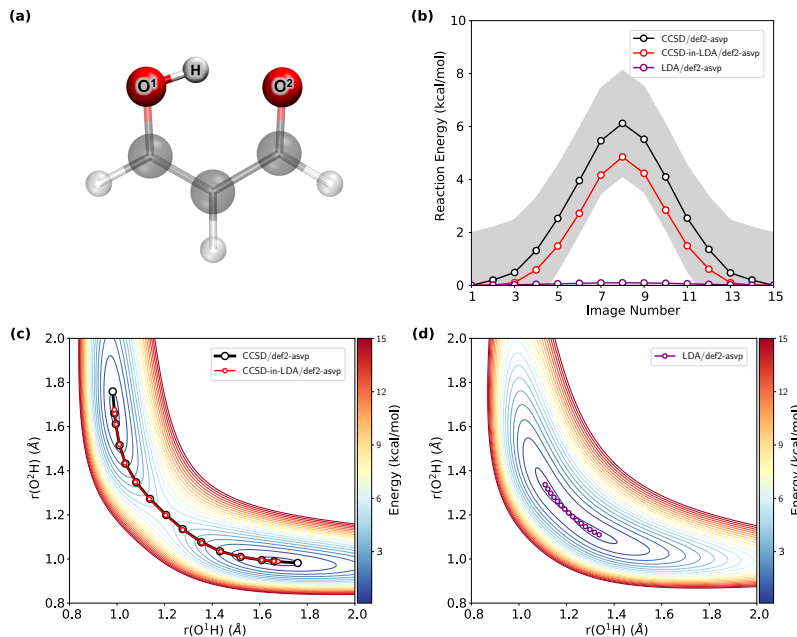


FIG. 3. (a) The ground state geometry of malondialdehyde evaluated at the CCSD-in-LDA/def2-aSVP level of theory. The solid atoms are included in subsystem A and the transparent atoms are included in subsystem B. (b) The reaction barrier heights for the minimum energy reaction pathways for LDA/def2-aSVP, CCSD-in-LDA/def2-aSVP and CCSD/def2-aSVP. (c), (d) Also shown are the minimum energy reaction pathways of the proton transfer in malondialdehyde as a function of the distance of the proton from the oxygen atoms,  $O^1$  on the  $x$ -axis,  $O^2$  on the  $y$ -axis for the CCSD-in-LDA/def2-aSVP and CCSD/def2-aSVP levels of theory, (c), and for the LDA/def2-aSVP level of theory, (d).

#### D. Cobalt-based Organometallic Complex: Minimum Energy Reaction Pathway

The minimum energy reaction pathway for the intramolecular proton transfer in a cobalt diimine-dioxime catalyst (Fig. 4a) is now investigated. Previously, the reaction pathway for the transfer of the  $[-NH]$  to form a cobalt hydride had been investigated using geometries obtained using DFT.<sup>20</sup> Fig. 4b shows the energy profile for this reaction determined by various levels of theory. We observe that the reaction pathway determined by the NEB optimized at the PBE0/cc-pVDZ level of theory (purple curve) predicts a barrier height of 5.45 kcal/mol. However, when single-point PNO-LCCSD-in-PBE0/cc-pVDZ embedding energy calculations are run on the PBE0 optimized geometries (blue curve), the barrier height is lowered to 3.35 kcal/mol and the position of the transition state is shifted towards the reactant. The NEB optimized at the CCSD-in-PBE0/cc-pVDZ embedding level of theory (red curve) shows an even lower barrier height of 2.61 kcal/mol and predicts a substantially different transition state geometry (Fig. 4c) than the DFT result. The difference between the transition states predicted by the PBE0/cc-pVDZ and PNO-LCCSD-in-PBE0/cc-pVDZ levels of theory is clearly seen in Fig. 4d, which shows the projection of the NEB onto the two dimensions of the Co-H and N-H bonds and with the position of the transition state geometry indicated with stars. This result clearly shows the large degree to which com-

monly employed DFT transition state geometries can differ from the CCSD-quality result that is obtained using projection-based embedding.

#### V. CONCLUSIONS

We present the derivation and numerical demonstration of analytical nuclear gradients for projection-based embedding both with and without AO truncation. A key aspect of the gradient theory is that the WF contributions can be evaluated using existing WF gradient implementations without the need for modification or additional programming, thereby allowing projection-based WF-in-DFT embedding gradients to be easily generalized to any combination of WF and KS-DFT methods. It is demonstrated that projection-based embedding gradients produce accurate geometries for a variety of benchmark systems, including for bond-lengths that span the interface between subsystems. Furthermore, in applications to both malondialdehyde and a transition-metal catalyst, WF-in-DFT minimum energy pathways obtained via the NEB method reveal large errors in DFT-computed transition-state energies and geometries. Finally, we note that the Lagrangian framework presented here can be used to derive other analytical gradients of the projection-based WF-in-DFT energy with respect to quantities such as electric and magnetic fields.

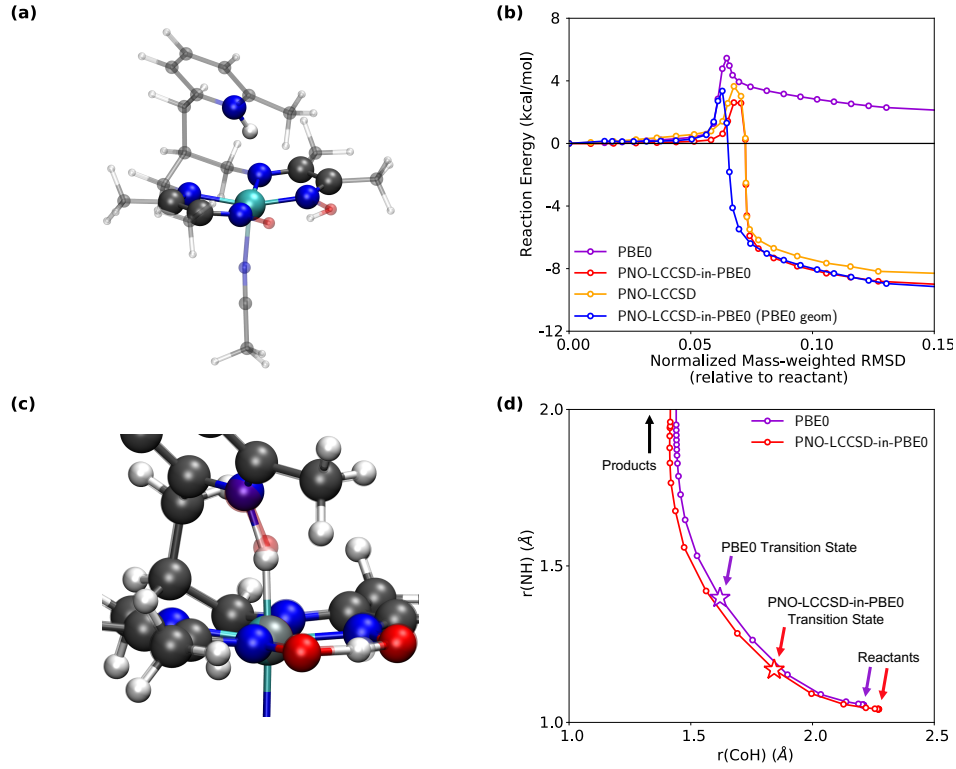


FIG. 4. All calculations used the cc-pVDZ basis set. (a) The optimized geometry for a cobalt-based organometallic complex calculated at the CCSD-in-PBE0 level of theory with AO truncation. The solid atoms are included in subsystem A while the transparent atoms are in subsystem B. (b) The minimum energy reaction pathway for PBE0, the reaction pathway for PNO-LCCSD-in-PBE0 using the PBE0 geometries, the reaction pathway for PNO-LCCSD-in-PBE0 using CCSD-in-PBE0 geometries, and the reaction pathway for PNO-LCCSD using the CCSD-in-PBE0 geometries. The  $x$ -axis is a coordinate constructed by taking a normalized mass-weighted RMSD of all images along the pathway with respect to the respective reactant and product. In comparing the purple and blue curves versus the orange and red curves, note that the transition state position in these normalized coordinates is affected by changes in the geometries of the reactant and product. (c) A zoomed-in picture of the transition state geometries predicted by PBE0 (opaque atoms) and PNO-LCCSD-in-PBE0 (transparent atoms) levels of theory. The proton placement between the nitrogen and cobalt center at the PNO-LCCSD-in-PBE0 level of theory is highlighted in red. (d) The minimum energy reaction pathways of the proton transfer for PNO-LCCSD-in-PBE0 and PBE0 as a function of the distance of the proton from the cobalt atom on the  $x$ -axis, and the nitrogen atom on the  $y$ -axis. The placement of the transition states are highlighted for each level of theory.

## ACKNOWLEDGMENTS

We thank Matthew Welborn for helpful discussions. This material is based upon work supported by the U.S. Army Research Laboratory under Grant No. W911NF-12-2-0023 (S.J.R.L.). S.J.R.L. thanks the Caltech Resnick Sustainability Institute for a graduate fellowship. T.F.M. and F.R.M. acknowledge joint support from the DOE (Award No. DEFOA-0001912), and F.R.M. acknowledges support from the Engineering and Physical Sciences Research Council for funding (EP/M013111/1).

## VI. SUPPORTING INFORMATION

All geometries used in all tables and figures are available for download.

## Appendix A: Pipek-Mezey Localization

Equation 14 from the main text

$$\begin{aligned} (\mathbf{a}[\mathbf{z}^{\text{loc}}])_{pi} &= \sum_{\mu} C_{\mu,p} \left( \sum_{k>l} z_{kl}^{\text{loc}} \frac{\partial r_{kl}}{\partial C_{\mu,q}} \right) \\ &= \sum_{k>l} \mathcal{B}_{pi,kl} z_{kl}^{\text{loc}} \end{aligned} \quad (\text{A1})$$

corresponds to the derivative of the localization conditions, Eq. 7, with respect to  $\mathbf{C}$ , where

$$\begin{aligned} \mathcal{B}_{pi,kl} &= \sum_C \left[ \left( 2S_{pk}^C \delta_{ki} - 2S_{pl}^C \delta_{li} \right) S_{kl}^C \right. \\ &\quad \left. + \left( S_{kk}^C - S_{ll}^C \right) \left( S_{pl}^C \delta_{ki} + S_{pk}^C \delta_{li} \right) \right]. \end{aligned} \quad (\text{A2})$$

Next, the overlap derivative contribution from the localization conditions from Eq. 34 is

$$\sum_{i>j} \frac{\partial r_{ij}}{\partial S_{\mu\nu}} z_{ij}^{\text{loc}} = \sum_{i>j} z_{ij}^{\text{loc}} (1 - \mathcal{P}_{ij}) \sum_C \left[ 2L_{\mu,i} L_{\nu,i} S_{ij}^C + S_{ii}^C (L_{\mu,i} L_{\nu,j} + L_{\mu,j} L_{\nu,i}) \right] \Big|_{\mu \in C}, \quad (\text{A3})$$

where  $\mathcal{P}_{ij}$  permutes the indices  $i$  and  $j$ , and  $\mu$  is restricted to atomic orbitals on atom  $C$ .

## Appendix B: Orbital Derivatives of Projection-based WF-in-DFT Embedding Energy

This appendix provides additional details for the terms in Eqs. 13 and 15 of the main text. The derivative of the projection-based WF-in-DFT embedding energy and the WF constraints with respect to the MO coefficients shown in Eq. 13 is

$$\begin{aligned} E_{pq} &= \sum_{\mu} C_{\mu,p} \left( \frac{\partial E_{\text{WF-in-DFT}}[\tilde{\Psi}^A; \gamma^A, \gamma^B]}{\partial C_{\mu,q}} + \frac{\partial}{\partial C_{\mu,q}} \sum_s \Lambda_s^{\text{WF,A}} c_s \right) \\ &= \sum_{\mu} C_{\mu,p} \left( \frac{\partial E_{\text{DFT}}[\gamma^A + \gamma^B]}{\partial C_{\mu,q}} - \frac{\partial E_{\text{DFT}}[\gamma^A]}{\partial C_{\mu,q}} + \text{tr} \left[ \frac{\partial (\tilde{\mathbf{d}}_{\text{rel}}^A - \gamma^A)}{\partial C_{\mu,q}} \mathbf{v}_{\text{emb}} \right] + \text{tr} \left[ (\tilde{\mathbf{d}}_{\text{rel}}^A - \gamma^A) \frac{\partial \mathbf{v}_{\text{emb}}}{\partial C_{\mu,q}} \right] \right. \\ &\quad \left. + \mu \text{tr} \left[ \frac{\partial \tilde{\mathbf{d}}_{\text{rel}}^A}{\partial C_{\mu,q}} \mathbf{P}^B \right] + \mu \text{tr} \left[ \tilde{\mathbf{d}}_{\text{rel}}^A \frac{\partial \mathbf{P}^B}{\partial C_{\mu,q}} \right] \right), \end{aligned} \quad (\text{B1})$$

where the partial derivative of the WF constraints causes the appearance of the WF relaxed density,  $\tilde{\mathbf{d}}_{\text{rel}}^A$ , in the last four terms on the RHS of Eq. B1. Equation B1 simplifies to

$$E_{pq} = 4(\mathbf{F}[\gamma^A + \gamma^B])_{pq} \Big|_{q \in \text{occ}} - 4(\mathbf{F}[\gamma^A])_{pq} \Big|_{q \in A} - 4(\mathbf{v}_{\text{emb}})_{pq} \Big|_{q \in A} + 4(\mathbf{M}[\tilde{\mathbf{d}}_{\text{rel}}^A - \gamma^A])_{pq}, \quad (\text{B2})$$

where  $q \in \text{occ}$  indicates that the index  $q$  is restricted to LMOs,  $q \in A$  indicates that  $q$  is restricted to LMOs in subsystem A, and  $\mathbf{F}$  is the KS Fock matrix evaluated with the bracketed density. The last term on the RHS of Eq. B2

$$(\mathbf{M}[\gamma])_{pq} = \frac{1}{4} \sum_{\mu} C_{\mu,p} \left( \text{tr} \left[ \gamma \frac{\partial \mathbf{v}_{\text{emb}}[\gamma^A, \gamma^B]}{\partial C_{\mu,q}} \right] + \mu \text{tr} \left[ \gamma \frac{\partial \mathbf{P}^B}{\partial C_{\mu,q}} \right] \right), \quad (\text{B3})$$

simplifies to

$$\begin{aligned} (\mathbf{M}[\gamma])_{pq} &= \sum_{\mu\nu} C_{\mu,p} \sum_{\lambda\sigma} \gamma_{\lambda\sigma} \left( (\mu\nu|\lambda\sigma) - \frac{1}{2} x_f(\mu\lambda|\nu\sigma) \right) L_{\nu,q} \Big|_{q \in B} + \mu (\mathbf{C}^\dagger \mathbf{P}[\gamma] \mathbf{L})_{pq} \Big|_{q \in B} \\ &\quad + (\tilde{\mathbf{v}}_{\text{xc}}[\gamma^A + \gamma^B, \gamma])_{pq} \Big|_{q \in \text{occ}} + (\tilde{\mathbf{v}}_{\text{xc}}[\gamma^A, \gamma])_{pq} \Big|_{q \in A} \end{aligned} \quad (\text{B4})$$

where

$$\mathbf{P}[\gamma] = \mathbf{S} \gamma \mathbf{S}. \quad (\text{B5})$$

In the current study, we employ both LDA and GGA exchange-correlation functionals; for the special case of LDA, the term  $\tilde{\mathbf{v}}_{\text{xc}}[\gamma^A + \gamma^B, \gamma]$  assumes the form

$$(\tilde{\mathbf{v}}_{\text{xc}}[\gamma^A + \gamma^B, \gamma])_{pq} = \sum_{mn} (pq|f_{\text{xc}}[\gamma^A + \gamma^B]|mn) \gamma_{mn}, \quad (\text{B6})$$

where  $f_{\text{xc}}$  is the XC kernel which is defined as the second derivative of the XC functional with respect to density.

The derivative of the Brillouin conditions in Eq. 15 can be expanded as follows.

$$\begin{aligned}
(\mathbf{D}[\mathbf{z}])_{pq} &= \sum_{ak} \mathcal{D}_{pq,ak} z_{ak} \\
&= \sum_{ak} z_{ak} \left[ (\mathbf{F}[\gamma^A + \gamma^B])_{pk} \delta_{aq} + (\mathbf{F}[\gamma^A + \gamma^B])_{ap} \delta_{kq} + 2 \sum_l \delta_{ql} \left( 2(ak|pl) - \frac{1}{2}x_f(ap|kl) - \frac{1}{2}x_f(al|kp) \right) \right. \\
&\quad \left. + \sum_{\mu\lambda\sigma} C_{\mu,p} C_{\lambda,a} \frac{\partial(\mathbf{v}_{xc}[\gamma^A + \gamma^B])_{\lambda\sigma}}{\partial C_{\mu,q}} L_{\sigma,k} \right] \\
&= (\mathbf{F}[\gamma^A + \gamma^B]\mathbf{z})_{pq} \Big|_{q \in \text{occ}} + (\mathbf{F}[\gamma^A + \gamma^B]\mathbf{z}^\dagger)_{pq} \Big|_{q \in \text{vir}} + 2(\mathbf{V}[\bar{\mathbf{z}}])_{pq} \Big|_{q \in \text{occ}},
\end{aligned} \tag{B7}$$

where  $\bar{\mathbf{z}} = \mathbf{z} + \mathbf{z}^\dagger$  and  $\mathbf{V}[\bar{\mathbf{z}}]$  is defined as

$$(\mathbf{V}[\bar{\mathbf{z}}])_{pq} \Big|_{q \in \text{occ}} = \sum_{mn} \bar{z}_{mn} \left( (mn|pq) - \frac{1}{2}x_f(mp|nq) \right) \Big|_{q \in \text{occ}} + (\tilde{\mathbf{v}}_{xc}[\gamma^A + \gamma^B, \bar{\mathbf{z}}])_{pq} \Big|_{q \in \text{occ}}. \tag{B8}$$

### Appendix C: Atomic Orbital Truncation

Projection-based WF-in-DFT embedding reduces the cost of the WF calculation on subsystem A by reducing the number of LMOs that are correlated at the WF level, but thus far leaves the virtual space untouched. However, the scaling of most WF methods is dominated by the number of virtual MOs (e.g.  $\mathcal{O}(v^4)$  for CCSD). One strategy has been to employ local correlation WF methods such as PNO-LMP2<sup>94</sup> and PNO-LCCSD<sup>95,96</sup> to describe subsystem A since these methods are able to leverage the reduced number of LMOs to significantly lower the number of occupied-virtual orbital pairs that need to be included, resulting in a cheap and accurate WF calculation. However, a real advantage of projection-based embedding hinges on being able to use any WF method to describe subsystem A. Therefore, having a more general approach to reduce the cost of the WF calculation on the subsystem A is desirable.

The AO truncation scheme devised by Bennie et al.<sup>18</sup> provides a simple way to significantly reduce the cost of the WF calculation by reducing the size of the basis used to describe subsystem A. The AOs that are discarded are selected through a single density threshold parameter: if the net Mulliken population, computed using the subsystem A density, of an AO is less than the specified threshold, it is removed from the basis set. This scheme has shown to greatly speedup up WF-in-DFT calculations at a small cost in accuracy in total and relative energies.<sup>18</sup> Additionally, it has the nice feature that given a fixed subsystem A, the size of the truncated subsystem A basis scales asymptotically as the size of the environment grows. This basis set modification does not cause any complications in the evaluation of the subsystem A WF gradient so existing implementations can be used without any modifications. The energy expression for a projection-based WF-in-DFT calculation with AO truncation using the so-called type-in-type correction<sup>18</sup>

is

$$\begin{aligned}
E_{\text{WF-in-DFT}}^{\text{trun}} [\tilde{\Psi}^{\text{A,trun}}; \tilde{\gamma}^{\text{A,trun}}; \gamma^{\text{A}}, \gamma^{\text{B}}] &= E_{\text{WF}} [\tilde{\Psi}^{\text{A,trun}}] \\
&- E_{\text{DFT}}^{\text{trun}} [\tilde{\gamma}^{\text{A,trun}}] + E_{\text{DFT}} [\gamma^{\text{A}} + \gamma^{\text{B}}] \\
&+ \text{tr} \left[ \left( \tilde{\mathbf{d}}^{\text{A,trun}} - \tilde{\gamma}^{\text{A,trun}} \right) \mathbf{v}_{\text{emb}}^{\text{trun}} [\gamma^{\text{A}}, \gamma^{\text{B}}] \right] \\
&+ \text{tr} \left[ \left( \tilde{\mathbf{d}}^{\text{A,trun}} - \tilde{\gamma}^{\text{A,trun}} \right) \mathbf{P}^{\text{B,trun}} \right]
\end{aligned} \tag{C1}$$

where  $\tilde{\Psi}^{\text{A,trun}}$  is the subsystem A WF in the truncated basis,  $\tilde{\gamma}^{\text{A,trun}}$  is the KS subsystem A one-particle density in the truncated basis,  $\gamma^{\text{A}}$  and  $\gamma^{\text{B}}$  are the KS subsystem A and B one-particle densities in the full basis respectively,  $\tilde{\mathbf{d}}^{\text{A,trun}}$  is the subsystem A one-particle reduced density matrix that corresponds to  $\tilde{\Psi}^{\text{A,trun}}$ ,  $\mathbf{v}_{\text{emb}}^{\text{trun}} [\gamma^{\text{A}}, \gamma^{\text{B}}]$  is the embedding potential in the truncated basis which is evaluated by

$$\mathbf{v}_{\text{emb}}^{\text{trun}} [\gamma^{\text{A}}, \gamma^{\text{B}}] = \mathbf{P}_t^\dagger \mathbf{v}_{\text{emb}} [\gamma^{\text{A}}, \gamma^{\text{B}}] \mathbf{P}_t, \tag{C2}$$

and  $\mathbf{P}^{\text{B,trun}}$  is the projection operator in the truncated basis which is evaluated by

$$\mathbf{P}^{\text{B,trun}} = \mathbf{P}_t^\dagger \mathbf{P}^{\text{B}} \mathbf{P}_t. \tag{C3}$$

Here,  $\mathbf{P}_t$  is the rectangular matrix that maps the full basis to the truncated basis which is created by starting with identity matrix and deleting columns corresponding to thrown away AO functions. We note that the even though the notation for Eq. C1 is different from the one used in Ref. 18 the approach is identical.

## Appendix D: Projection-based WF-in-DFT Gradient Theory with AO Truncation

### 1. Total Energy Lagrangian

We now derive the total energy Lagrangian for projection-based WF-in-DFT embedding with AO truncation. The WF-in-DFT AO truncation Lagrangian is

$$\begin{aligned} \mathcal{L} = & E_{\text{WF-in-DFT}}^{\text{trun}} [\tilde{\Psi}^{\text{A,trun}}; \tilde{\gamma}^{\text{A,trun}}; \gamma^{\text{A}}, \gamma^{\text{B}}] + \sum_s \Lambda_s^{\text{WF,A}} c_s \\ & - \sum_{ij \in \bar{\text{A}}} \bar{\epsilon}_{ij}^{\text{A}} [\bar{\mathbf{C}}^{\text{A}\dagger} \mathbf{S} \bar{\mathbf{C}}^{\text{A}} - \mathbf{1}]_{ij} \\ & + \sum_{pq} x_{pq} [\mathbf{C}^{\dagger} \mathbf{S} \mathbf{C} - \mathbf{1}]_{pq} + \sum_{i>j} z_{ij}^{\text{loc}} r_{ij} \\ & + \sum_{ai} z_{ai} (\mathbf{F}[\gamma^{\text{A}} + \gamma^{\text{B}}])_{ai}, \end{aligned} \quad (\text{D1})$$

where the bar superscript refers to subsystem A quantities optimized by the KS functional in the truncated basis. The constraints that appear in Eq. D1 are all the same as those that appear in Eq. 6 from the main text, except for the third term on the RHS of Eq. D1. This term constrains the MOs,  $\bar{\mathbf{C}}^{\text{A}}$ , to be orthogonal.

#### a. Minimizing the Lagrangian with respect to the variational parameters of the WF method – $\tilde{\Psi}^{\text{A,trun}}$ .

Minimizing the WF-in-DFT AO truncation Lagrangian with respect to  $\tilde{\Psi}^{\text{A,trun}}$  simplifies to the minimization of the subsystem A WF energy and the WF constraints (as explained in section IID), which corresponds to the conventional WF Lagrangian used to derive WF gradient theories, albeit in the truncated basis.

#### b. Minimizing the Lagrangian with respect to the MO coefficients, $\bar{\mathbf{C}}^{\text{A}}$ .

The minimization of the Lagrangian with respect to the optimized KS MO coefficients in the truncated basis,  $\bar{\mathbf{C}}^{\text{A}}$ , results in the SCF equations using the embedded Fock matrix.

$$\begin{aligned} \sum_{\mu} \bar{C}_{\mu,i}^{\text{A}} \left( \frac{\partial E_{\text{DFT}}^{\text{trun}}[\tilde{\gamma}^{\text{A,trun}}]}{\partial \bar{C}_{\mu,j}^{\text{A}}} + \frac{\partial \text{tr}[\tilde{\gamma}^{\text{A,trun}} \mathbf{v}_{\text{emb}}^{\text{trun}}[\gamma^{\text{A}}, \gamma^{\text{B}}]]}{\partial \bar{C}_{\mu,j}^{\text{A}}} \right. \\ \left. + \mu \frac{\partial \text{tr}[\tilde{\gamma}^{\text{A,trun}} \mathbf{P}^{\text{B,trun}}]}{\partial \bar{C}_{\mu,j}^{\text{A}}} \right) = \\ = \sum_{\mu} \bar{C}_{\mu,i}^{\text{A}} \sum_{kl \in \bar{\text{A}}} \bar{\epsilon}_{kl}^{\text{A}} \frac{\partial \bar{S}_{kl}^{\text{A}}}{\partial \bar{C}_{\mu,j}^{\text{A}}} \end{aligned} \quad (\text{D2})$$

$$\begin{aligned} \sum_{\mu\nu} \bar{C}_{\mu,i}^{\text{A}} \left( (\mathbf{F}[\tilde{\gamma}^{\text{A,trun}}])_{\mu\nu} + (\mathbf{v}_{\text{emb}}^{\text{trun}}[\gamma^{\text{A}}, \gamma^{\text{B}}])_{\mu\nu} \right. \\ \left. + \mu [\mathbf{P}^{\text{B,trun}}]_{\mu\nu} \right) \bar{C}_{\nu,j}^{\text{A}} = \frac{1}{2} \bar{\epsilon}_{ij}^{\text{A}} \Big|_{ij \in \text{A}} \end{aligned} \quad (\text{D3})$$

$$(\mathbf{F}^{\text{A}}[\tilde{\gamma}^{\text{A,trun}}])_{ij} \Big|_{ij \in \text{A}} = \frac{1}{2} \bar{\epsilon}_{ij}^{\text{A}} \Big|_{ij \in \text{A}} \quad (\text{D4})$$

Therefore, the Lagrange multipliers  $\frac{1}{2} \bar{\epsilon}^{\text{A}}$  are simply the MO eigen energies of the KS optimized subsystem A MOs in the truncated basis.

#### c. Minimize the Lagrangian with respect to the MO coefficients, $\mathbf{C}$ .

The minimization of the Lagrangian with respect to the KS MO coefficients in the full basis,  $\mathbf{C}$ , is

$$\begin{aligned} \sum_{\mu} C_{\mu,p} \frac{\partial \mathcal{L}}{\partial C_{\mu,q}} = E_{pq} + (\mathbf{a}[\mathbf{z}^{\text{loc}}])_{pi} \\ + (\mathbf{D}[\mathbf{z}])_{pq} + 2x_{pq} = 0. \end{aligned} \quad (\text{D5})$$

where only the matrix  $\mathbf{E}$  differs from the ones outlined in Eqns. 13-16.

$$\begin{aligned} E_{pq} = 4(\mathbf{F}[\gamma^{\text{A}} + \gamma^{\text{B}}])_{pq} \Big|_{q \in \text{occ}} \\ + 4 \left( \mathbf{M}[\mathbf{P}_t(\tilde{\mathbf{d}}_{\text{rel}}^{\text{A,trun}} - \tilde{\gamma}^{\text{A,trun}}) \mathbf{P}_t^{\dagger}] \right)_{pq} \end{aligned} \quad (\text{D6})$$

With the updated  $\mathbf{E}$  matrix, the Lagrangian multipliers are solved in the same way as outlined for the WF-in-DFT Lagrangian multipliers.

#### d. Gradient of the Total Energy

Once the Lagrangian is minimized with respect to all variational parameters, the gradient of the energy with respect to nuclear coordinate,  $q$ , takes the form

$$\begin{aligned} E_{\text{WF-in-DFT}}^{\text{trun},(q)} = E_{\text{WF}}^{\text{trun},q}[\tilde{\Psi}^{\text{A,trun}}] - E_{\text{DFT}}^{\text{trun},q}[\tilde{\gamma}^{\text{A,trun}}] \\ + \text{tr}[\mathbf{d}_a \mathbf{h}^{(q)}] + \text{tr}[\mathbf{X} \mathbf{S}^{(q)}] + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} D_{\mu\nu\lambda\sigma} (\mu\nu|\lambda\sigma)^{(q)} \\ + E_{\text{xc}}^{(q)}[\gamma^{\text{A}} + \gamma^{\text{B}}] + \text{tr}[\mathbf{d}_c(\mathbf{v}_{\text{xc}}^{(q)}[\gamma^{\text{A}} + \gamma^{\text{B}}] - \mathbf{v}_{\text{xc}}^{(q)}[\gamma^{\text{A}}])], \end{aligned} \quad (\text{D7})$$

where the first two terms on the RHS of Eq. D7,  $E_{\text{WF}}^{\text{trun},q}$  and  $E_{\text{DFT}}^{\text{trun},q}$ , are the total derivative of the truncated subsystem A WF and KS energy respectively, minus the embedding contribution,  $\mathbf{v}_{\text{emb}}^{\text{trun},(q)}$ . These two terms are calculated using existing gradient implementations, whereas the embedding contribution has been folded into the remaining terms. The effective one-particle densities  $\mathbf{d}_a$ ,  $\mathbf{d}_b$  and  $\mathbf{d}_c$  are

$$\mathbf{d}_a = \gamma^{\text{A}} + \gamma^{\text{B}} + \mathbf{C} \mathbf{z} \mathbf{C}^{\dagger}, \quad (\text{D8})$$

$$\mathbf{d}_b = \gamma^A + \gamma^B + 2\mathbf{CzC}^\dagger + 2\mathbf{d}_c, \quad (\text{D9})$$

and

$$\mathbf{d}_c = \mathbf{P}_t \left( \tilde{\mathbf{d}}_{\text{rel}}^{\text{A, trun}} - \bar{\gamma}^{\text{A, trun}} \right) \mathbf{P}_t^\dagger. \quad (\text{D10})$$

The effective two-particle density  $\mathbf{D}$  is

$$\begin{aligned} D_{\mu\nu\lambda\sigma} &= (\gamma^A + \gamma^B)_{\mu\nu} (\mathbf{d}_b)_{\lambda\sigma} - 2\gamma_{\mu\nu}^A (\mathbf{d}_c)_{\lambda\sigma} \\ &\quad - \frac{1}{2} x_f \left( (\gamma^A + \gamma^B)_{\mu\lambda} (\mathbf{d}_b)_{\nu\sigma} - 2\gamma_{\mu\lambda}^A (\mathbf{d}_c)_{\nu\sigma} \right). \end{aligned} \quad (\text{D11})$$

The matrix  $\mathbf{X}$  is

$$\begin{aligned} \mathbf{X} &= \mathbf{X}^{\text{loc}} - \frac{1}{2} \mathbf{L} (\mathbf{E} + 2\mathbf{V}[\bar{\mathbf{z}}]) \mathbf{L}^\dagger \\ &\quad - \frac{1}{2} \left( \mathbf{C}_v(\mathbf{zF}) \mathbf{L}^\dagger + (\mathbf{C}_v(\mathbf{zF}) \mathbf{L}^\dagger)^\dagger \right) \\ &\quad + \mu \left( \mathbf{d}_c \mathbf{S} \gamma^B + \gamma^B \mathbf{S} \mathbf{d}_c \right), \end{aligned} \quad (\text{D12})$$

where

$$(\mathbf{X}^{\text{loc}})_{\mu\nu} = -\frac{1}{2} (\mathbf{L} \mathbf{a} [\mathbf{z}^{\text{loc}}] \mathbf{L}^\dagger)_{\mu\nu} + \sum_{i>j} \frac{\partial r_{ij}}{\partial S_{\mu\nu}} z_{ij}^{\text{loc}}. \quad (\text{D13})$$

- <sup>1</sup>K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, *Chemical Physics Letters* **313**, 701 (1999).
- <sup>2</sup>V. Deev and M. A. Collins, *The Journal of Chemical Physics* **122**, 154102 (2005).
- <sup>3</sup>M. A. Collins and V. A. Deev, *Journal of Chemical Physics* **125** (2006), 10.1063/1.2347710.
- <sup>4</sup>D. G. Fedorov and K. Kitaura, *Journal of Physical Chemistry A* **111**, 6904 (2007).
- <sup>5</sup>P. Elliott, M. H. Cohen, A. Wasserman, and K. Burke, *Journal of Chemical Theory and Computation* **5**, 827 (2009).
- <sup>6</sup>J. D. Goodpaster, N. Ananth, F. R. Manby, and T. F. Miller III, *The Journal of Chemical Physics* **133**, 084103 (2010).
- <sup>7</sup>C. Huang and E. A. Carter, *The Journal of Chemical Physics* **135**, 194104 (2011).
- <sup>8</sup>J. D. Goodpaster, T. a. Barnes, and T. F. Miller III, *The Journal of Chemical Physics* **134**, 164108 (2011).
- <sup>9</sup>F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller III, *Journal of Chemical Theory and Computation* **8**, 2564 (2012).
- <sup>10</sup>J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, *The Journal of Chemical Physics* **137**, 224113 (2012).
- <sup>11</sup>M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, *Chemical Reviews* **112**, 632 (2012).
- <sup>12</sup>G. Knizia and G. K.-L. Chan, *Physical Review Letters* **109**, 186404 (2012).
- <sup>13</sup>T. A. Barnes, J. D. Goodpaster, F. R. Manby, and T. F. Miller III, *The Journal of Chemical Physics* **139**, 024103 (2013).
- <sup>14</sup>J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, *The Journal of Chemical Physics* **140**, 18A507 (2014).
- <sup>15</sup>D. Neuhauser, R. Baer, and E. Rabani, *The Journal of Chemical Physics* **141**, 041102 (2014).
- <sup>16</sup>T. A. Barnes, J. W. Kaminski, O. Borodin, and T. F. Miller III, *The Journal of Physical Chemistry C* **119**, 3865 (2015).
- <sup>17</sup>M. E. Fornace, J. Lee, K. Miyamoto, F. R. Manby, and T. F. Miller III, *Journal of Chemical Theory and Computation* **11**, 568 (2015).
- <sup>18</sup>S. J. Bennie, M. Stella, T. F. Miller III, and F. R. Manby, *The Journal of Chemical Physics* **143**, 024105 (2015).
- <sup>19</sup>M. Stella, S. J. Bennie, and F. R. Manby, *Molecular Physics* **113**, 1858 (2015).
- <sup>20</sup>P. Huo, C. Uyeda, J. D. Goodpaster, J. C. Peters, and T. F. Miller III, *ACS Catalysis* **6**, 6114 (2016).
- <sup>21</sup>S. J. Bennie, M. W. van der Kamp, R. C. R. Pennifold, M. Stella, F. R. Manby, and A. J. Mulholland, *Journal of Chemical Theory and Computation* **12**, 2689 (2016).
- <sup>22</sup>R. C. R. Pennifold, S. J. Bennie, T. F. Miller III, and F. R. Manby, *The Journal of Chemical Physics* **146**, 084113 (2017).
- <sup>23</sup>X. Zhang, S. J. Bennie, M. W. van der Kamp, D. R. Glowacki, F. R. Manby, and A. J. Mulholland, *Royal Society Open Science* **5**, 171390 (2018).
- <sup>24</sup>A. Chapovetsky, M. Welborn, J. M. Luna, R. Haiges, T. F. Miller III, and S. C. Marinescu, *ACS Central Science* **4**, 397 (2018).
- <sup>25</sup>A. H. Mühlbach and M. Reiher, *The Journal of Chemical Physics* **149**, 184104 (2018).
- <sup>26</sup>S. J. R. Lee, M. Welborn, F. R. Manby, and T. F. Miller III, *Accounts of Chemical Research* **52**, 1359 (2019).
- <sup>27</sup>M. Welborn, F. R. Manby, and T. F. Miller III, *The Journal of Chemical Physics* **149**, 144101 (2018).
- <sup>28</sup>A. P. de Lima Batista, A. G. S. de Oliveira-Filho, and S. E. Galembeck, *Physical Chemistry Chemical Physics* **19**, 13860 (2017).
- <sup>29</sup>S. J. Bennie, B. F. E. Curchod, F. R. Manby, and D. R. Glowacki, *The Journal of Physical Chemistry Letters* **8**, 5559 (2017).
- <sup>30</sup>X. Chen and C. Franklin Goldsmith, *Proceedings of the Combustion Institute* **37**, 3167 (2019).
- <sup>31</sup>R. M. Parrish, J. F. Gonthier, C. Corminbœuf, and C. D. Sherill, *The Journal of Chemical Physics* **143**, 051103 (2015).
- <sup>32</sup>F. Libisch, M. Marsman, J. Burgdörfer, and G. Kresse, *The Journal of Chemical Physics* **147**, 034110 (2017).
- <sup>33</sup>K. Yao, J. E. Herr, S. N. Brown, and J. Parkhill, *The Journal of Physical Chemistry Letters* **8**, 2689 (2017).
- <sup>34</sup>O. R. Meitei and A. Heßelmann, *Journal of Computational Chemistry* **38**, 2500 (2017).
- <sup>35</sup>D. V. Chulhai and J. D. Goodpaster, *Journal of Chemical Theory and Computation* **14**, 1928 (2018).
- <sup>36</sup>L. Lin and L. Zepeda-Núñez, *arXiv e-prints*, arXiv:1807.08859 (2018).
- <sup>37</sup>M. Böckers and J. Neugebauer, *The Journal of Chemical Physics* **149**, 074102 (2018).
- <sup>38</sup>A. Hesselmann and O. R. Meitei, *Journal of Theoretical and Computational Chemistry* **17**, 1850037 (2018).
- <sup>39</sup>T. Nagata, D. G. Fedorov, and K. Kitaura, *Chemical Physics Letters* **544**, 87 (2012).
- <sup>40</sup>H. Nakata, T. Nagata, D. G. Fedorov, S. Yokojima, K. Kitaura, and S. Nakamura, *The Journal of Chemical Physics* **138**, 164103 (2013).
- <sup>41</sup>K. R. Brorsen, F. Zahariev, H. Nakata, D. G. Fedorov, and M. S. Gordon, *Journal of Chemical Theory and Computation* **10**, 5297 (2014).
- <sup>42</sup>A. Warshel and M. Levitt, *Journal of Molecular Biology* **103**, 227 (1976).
- <sup>43</sup>M. J. Field, P. A. Bash, and M. Karplus, *Journal of Computational Chemistry* **11**, 700 (1990).
- <sup>44</sup>H. Lin and D. G. Truhlar, *Theoretical Chemistry Accounts* **117**, 185 (2007).
- <sup>45</sup>S. Dapprich, I. Komáromi, K. Byun, K. Morokuma, and M. J. Frisch, *Journal of Molecular Structure: THEOCHEM* **461-462**, 1 (1999).
- <sup>46</sup>H. P. Hratchian, P. V. Parandekar, K. Raghavachari, M. J. Frisch, and T. Vreven, *The Journal of Chemical Physics* **128**, 034107 (2008).
- <sup>47</sup>N. J. Mayhall, K. Raghavachari, and H. P. Hratchian, *Journal of Chemical Physics* **132**, 114107 (2010).
- <sup>48</sup>H. P. Hratchian, A. V. Krukau, P. V. Parandekar, M. J. Frisch, and K. Raghavachari, *The Journal of Chemical Physics* **135**, 014105 (2011).
- <sup>49</sup>F. R. Manby, T. F. Miller III, P. J. Bygrave, F. Ding, T. Dresselhaus, F. A. Batista-Romero, A. Buccheri, C. Bungey, S. J. R. Lee, R. Meli, K. Miyamoto, C. Steinmann, T. Tsuchiya, M. Welborn, T. Wiles, and Z. Williams, (2019), 10.26434/chem-

- rxiv.7762646.v2.
- <sup>50</sup>M. Dulak, J. W. Kamiński, and T. A. Wesolowski, *Journal of Chemical Theory and Computation* **3**, 735 (2007).
  - <sup>51</sup>J. Heuser and S. Höfener, *Journal of Computational Chemistry* **37**, 1092 (2016).
  - <sup>52</sup>J. Heuser and S. Höfener, *Journal of Computational Chemistry* **38**, 2316 (2017).
  - <sup>53</sup>J. Heuser and S. Höfener, *Journal of Chemical Theory and Computation* **14**, 4616 (2018).
  - <sup>54</sup>A. Kovyrshin and J. Neugebauer, *Physical Chemistry Chemical Physics* **18**, 20955 (2016).
  - <sup>55</sup>D. Schlüns, M. Franchini, A. W. Götz, J. Neugebauer, C. R. Jacob, and L. Visscher, *Journal of Computational Chemistry* **38**, 238 (2017).
  - <sup>56</sup>K. Klahr, D. Schlüns, and J. Neugebauer, *Journal of Chemical Theory and Computation* **14**, 5631 (2018).
  - <sup>57</sup>H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györfy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. J. Bennie, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, S. J. R. Lee, Y. Liu, A. W. Lloyd, Q. Ma, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, T. F. Müller III, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and M. Welborn, "Molpro, version 2019.1, a package of ab initio programs," (2019), see <http://www.molpro.net>.
  - <sup>58</sup>T. A. Wesolowski and A. Warshel, *The Journal of Physical Chemistry* **97**, 8050 (1993).
  - <sup>59</sup>A. W. Gotz, S. M. Beyhan, and L. Visscher, *Journal of Chemical Theory and Computation* **5**, 3161 (2009).
  - <sup>60</sup>P. Pulay, *Molecular Physics* **17**, 197 (1969).
  - <sup>61</sup>J. Pipek and P. G. Mezey, *The Journal of Chemical Physics* **90**, 4916 (1989).
  - <sup>62</sup>J. M. Foster and S. F. Boys, *Reviews of Modern Physics* **32**, 300 (1960).
  - <sup>63</sup>G. Knizia, *Journal of Chemical Theory and Computation* **9**, 4834 (2013).
  - <sup>64</sup>M. Schütz, H.-J. Werner, R. Lindh, and F. R. Manby, *The Journal of Chemical Physics* **121**, 737 (2004).
  - <sup>65</sup>B. Hégyel, P. R. Nagy, G. G. Ferenczy, and M. Kállay, *The Journal of Chemical Physics* **145**, 064107 (2016).
  - <sup>66</sup>P. Hohenberg and W. Kohn, *Physical Review* **136**, B864 (1964).
  - <sup>67</sup>S. H. Vosko, L. Wilk, and M. Nusair, *Canadian Journal of Physics* **58**, 1200 (1980).
  - <sup>68</sup>J. P. Perdew, K. Burke, and M. Ernzerhof, *Physical Review Letters* **77**, 3865 (1996).
  - <sup>69</sup>C. Adamo and V. Barone, *The Journal of Chemical Physics* **110**, 6158 (1999).
  - <sup>70</sup>F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).
  - <sup>71</sup>F. Weigend, *Physical Chemistry Chemical Physics* **8**, 1057 (2006).
  - <sup>72</sup>T. H. Dunning, *The Journal of Chemical Physics* **90**, 1007 (1989).
  - <sup>73</sup>W. J. Hehre, R. Ditchfield, and J. A. Pople, *The Journal of Chemical Physics* **56**, 2257 (1972).
  - <sup>74</sup>F. Neese, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **8**, e1327 (2018).
  - <sup>75</sup>C. Møller and M. S. Plesset, *Physical Review* **46**, 618 (1934).
  - <sup>76</sup>G. E. Scuseria, C. L. Janssen, and H. F. Schaefer, *The Journal of Chemical Physics* **89**, 7382 (1988).
  - <sup>77</sup>G. D. Purvis and R. J. Bartlett, *The Journal of Chemical Physics* **76**, 1910 (1982).
  - <sup>78</sup>K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, *Chemical Physics Letters* **157**, 479 (1989).
  - <sup>79</sup>U. Bozkaya and C. D. Sherrill, *The Journal of Chemical Physics* **144**, 174103 (2016).
  - <sup>80</sup>U. Bozkaya and C. D. Sherrill, *The Journal of Chemical Physics* **147**, 044104 (2017).
  - <sup>81</sup>E. R. Davidson, *Journal of Computational Physics* **17**, 87 (1975).
  - <sup>82</sup>J. Kauczor, P. Jørgensen, and P. Norman, *Journal of Chemical Theory and Computation* **7**, 1610 (2011).
  - <sup>83</sup>L.-P. Wang and C. Song, *The Journal of Chemical Physics* **144**, 214108 (2016).
  - <sup>84</sup>L.-P. Wang, "Geometric," <https://github.com/leeping/geometric> (2019).
  - <sup>85</sup>G. Henkelman and H. Jónsson, *The Journal of Chemical Physics* **113**, 9978 (2000).
  - <sup>86</sup>A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dulak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, *Journal of Physics: Condensed Matter* **29**, 273002 (2017).
  - <sup>87</sup>S. Smidstrup, A. Pedersen, K. Stokbro, and H. Jónsson, *The Journal of Chemical Physics* **140**, 214106 (2014).
  - <sup>88</sup>G. Henkelman, B. P. Uberuaga, and H. Jónsson, *The Journal of Chemical Physics* **113**, 9901 (2000).
  - <sup>89</sup>E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, *Physical Review Letters* **97**, 170201 (2006).
  - <sup>90</sup>M. Schwilk, D. Usvyat, and H.-J. Werner, *The Journal of Chemical Physics* **142**, 121102 (2015).
  - <sup>91</sup>F. Weigend, *Physical Chemistry Chemical Physics* **4**, 4285 (2002).
  - <sup>92</sup>F. Weigend, *Journal of Computational Chemistry* **29**, 167 (2008).
  - <sup>93</sup>F. Weigend, A. Köhn, and C. Hättig, *The Journal of Chemical Physics* **116**, 3175 (2002).
  - <sup>94</sup>H.-J. Werner, G. Knizia, C. Krause, M. Schwilk, and M. Dornbach, *Journal of Chemical Theory and Computation* **11**, 484 (2015).
  - <sup>95</sup>M. Schwilk, Q. Ma, C. Köppl, and H.-J. Werner, *Journal of Chemical Theory and Computation* **13**, 3650 (2017).
  - <sup>96</sup>Q. Ma and H.-J. Werner, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **8**, e1371 (2018).